ChinaXiv合作期刊 图书情报工作
第 64 卷 第 21 期 2020 年 11 月

弱引文关系视角下跨学科相关知识组合识别方法探讨*
——以情报学为例

■ 牌艳欣　李长玲　徐璐
山东理工大学科技信息研究所　淄博 255049

摘　要：[目的/意义]科学系统的复杂化，使跨学科合作成为现代科学创新研究的重要范式和必然趋势。识别具有高度合作潜力的跨学科相关知识组合，成为促进跨学科科研合作创新的关键。[方法/过程]首先，选择目标学科源文献及其跨学科参考文献、跨学科引证文献，构建基于关键词的跨学科知识弱引文关联网络；其次，划分知识媒介 b 的类型，并识别目标学科知识节点 a－知识媒介 b－跨学科相关知识 c 的弱关系结构；最后，定义目标学科知识节点影响力指数 A_1、知识媒介影响力指数 B_1、跨学科知识相关性指数 C_1、跨学科知识组合 a－c 潜在合作指数 P，识别合作潜力值高的跨学科相关知识组合。[结果/结论]选择情报学领域 9 种 CSSCI 期刊 2015－2019 年的载文及其跨学科参考与引证文献为样本进行实证研究，验证基于弱引文关系的跨学科相关知识组合发现方法的有效性与可行性，并识别得到"科研合作"－"知识流动"－"种群动力学模型"等情报学科的高合作潜力的跨学科相关知识组合。

关键词：弱关系　引文分析　a-b-c 弱连接　跨学科相关知识组合　潜在合作指数 P
分类号：G250
DOI：10.13266/j.issn.0252－3116.2020.21.014

1　引言

学科是人类对知识的系统归类。科学系统的逐渐复杂化，致使诸多社会问题和科学研究都无法依靠单一学科知识解决，打破学科界限的科学知识交流与合作日益频繁，跨学科研究成为现代科学创新发展不可或缺的模式[1]。在此背景下，为解决本学科研究难题、突破科研瓶颈或实现科研创新，需要不断获取并动态引入其他学科的相关概念、理论、方法和技术等，进行跨学科合作研究。然而在实际工作中，研究者一般对本学科知识相对了解，对哪些跨学科知识可以进行合作研究，没有基本的把握。因此，跨学科相关知识组合的识别，成为跨学科合作研究的关键。

"弱关系"（strength of weak ties）起源于社会学研究领域，释义为两个行动者之间的短暂社会接触[2]。其提出者 M. Granovetter 教授[2]指出强关系使组织内部联系密切且稳定，而弱关系则为不同群体、组织之间的信息交流提供了重要途径，使相互孤立的各子群之间开始建立联系，并且伴随这种弱联系的不断加强，不同信息交流的范围进一步扩大，加速了信息的传播、融合、发展及创新。随后学者 J. P. Onnela 等[3]、E. Bakshy 等[4]、J. Zhao 等[5]、E. David 等[6]同样在研究中论述了这一观点。S. K. Genius[7]研究结果表明：与强连接关系相比，弱关系能传递更多潜在的、多样化的、非冗余的知识资源。A. Abbasi 等[8]、M. Bettoni 等[9]、L. Y. Yang 等[10]发现弱连接关系更易使个体与知识网络中其他子网中的个体建立更为广泛的关联关系，对促进科研合作具有积极作用。

目前，在科学计量学领域有学者用共词网络中的弱关系进行相关研究。隗玲等[11]基于弱关系理论构建高频词弱共现网络，根据主题之间弱连接的特点分析学科交叉的模式，探讨情报学学科微观层面的交叉性。李长玲等[12]、刘小慧等[13]利用开放式、闭合式的非相关知识发现方法，识别情报学与计算机科学领域的潜在跨学科合作研究主题。但 M. Song 等[14]对比发现，引文比共现能识别数量更多的关系配对，且配对关

* 本文系国家社会科学基金重点项目"跨学科潜在知识生长点识别与创新趋势预测研究"（项目编号：19ATQ006）研究成果之一。
作者简介：牌艳欣（ORCID：0000－0001－6266－4820），硕士研究生；李长玲（ORCID：0000－0001－6266－4820），教授，硕士生导师，通讯作者，E-mail：lichl69@163.com；徐璐（ORCID：0000－0002－6086－464X），硕士研究生。
收稿日期：2020－06－03　修回日期：2020－07－21　本文起止页码：111－119　本文责任编辑：易飞

系更具独特性和多样性。

引文是有效揭示不同学科之间知识扩散、关联与演化路径最为直接的表达，有效促进了跨学科知识的协同、交叉、融合、发展与创新[15]。因此，跨学科引文分析是跨学科知识流动的有效识别工具[16]。引文网络中存在知识的弱关联，根据弱关系的定义，可以分为以下几种类型：①低阈值关系节点。网络中节点关系强度低于阈值、与强关系对立的一类关系连接为弱关系[11]。例如源文献与其参考或引证文献形成的引文知识网络中连接强度较小的节点关系。②子网间关系节点。不同子网间节点联系相对稀疏，则为弱关系[17]。例如跨学科引用/被引用网络中不同学科知识之间的引用/被引用关系节点。③间接联系关系节点。通过其他节点建立间接联系的非相关知识，具有弱关系。例如共引或共被引关系节点，或有较好的引用关系但没有共现关系的知识节点。目前很少有学者利用引文网络中存在的弱关系进行知识挖掘、发现的研究，更少有涉及跨学科的研究，杜德慧等[18]以情报学科源文献关键词－跨学科参考文献关键词作为关联数据，构建跨学科知识引文网络，去除关键词共现关系，形成弱关系引用网络，用于识别跨学科相关知识。

在引文网络的低阈值、子网间、间接联系 3 种弱关系中，没有直接联系的子网间的间接关系节点，更具有潜在合作价值，也更适合于跨学科相关知识发现[14]。因此，本文提取学科源文献关键词及其跨学科参考文献关键词、其跨学科引证文献关键词，构建跨学科知识关联网络，在子网间识别间接联系知识节点，作为弱关系关联数据，探讨跨学科相关知识组合的发现方法；并以情报学为例进行实证研究，以期为该学科未来开展跨学科合作研究，进行有针对性的科研创新，提供重要的决策参考。

2　研究步骤与方法

本文基于跨学科子网间、间接联系知识节点的弱引文关系，识别跨学科相关知识组合。步骤分为以下 3 个阶段：①构建跨学科知识引用/被引用弱关联网络；②识别跨学科引文网络中的目标学科知识节点 a－知识媒介 b－跨学科知识 c 的弱关系结构；③构建跨学科相关知识组合 a-c 评价识别模型。具体步骤与方法如图 1 所示：

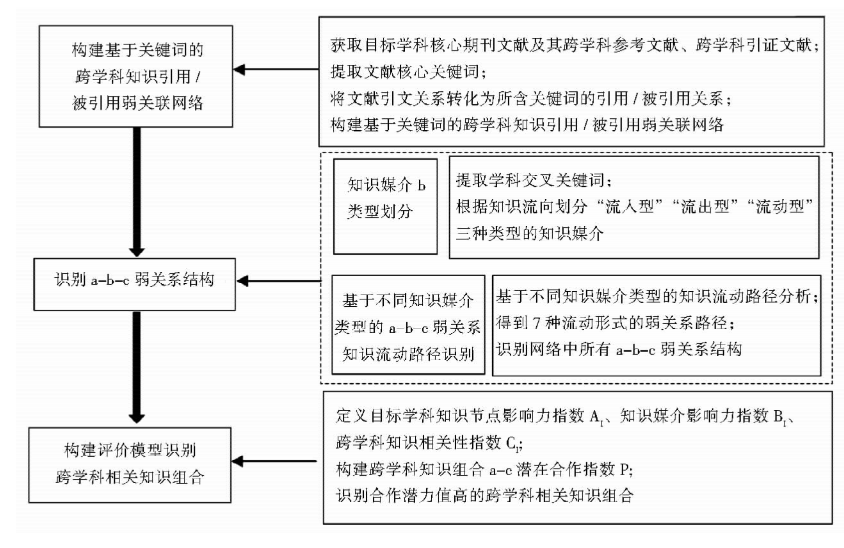


图 1　研究框架

2.1　构建跨学科知识引用/被引用弱关联网络

所研究学科，即目标学科。目标学科源文献的参考与引证文献，一部分来源于目标学科，即本学科；另一部分来源于非目标学科，即跨学科。知识以文献作为载体，通过文献间的引用与被引用实现不同学科知识流动与融合，构成知识动态交流与关联网络[19]。这一网络是由若干强连接关系和弱连接关系构成的交互网络[20]。其中，目标学科文献之间基于引用与被引用产生的多为内部知识交流，知识关联密切、共享性强，能够进行深层次交流，属于强关系连接；而跨学科参考/引证文献与目标学科知识的交流频次少，且关联程度低，则为弱连接关系。利用目标学科与跨学科关键词构建引文知识关联网络 G，见图 2。

2.2　识别跨学科引文网络中的 a-b-c 弱关系结构

弱关系理论指出，在社交网络中，若 a 和 c 有一个共同的朋友 b，那么 a 和 c 成为朋友的概率会增加，从而建立某种联系[2]。同样在跨学科知识引文网络中也适用该原理，见图 3。

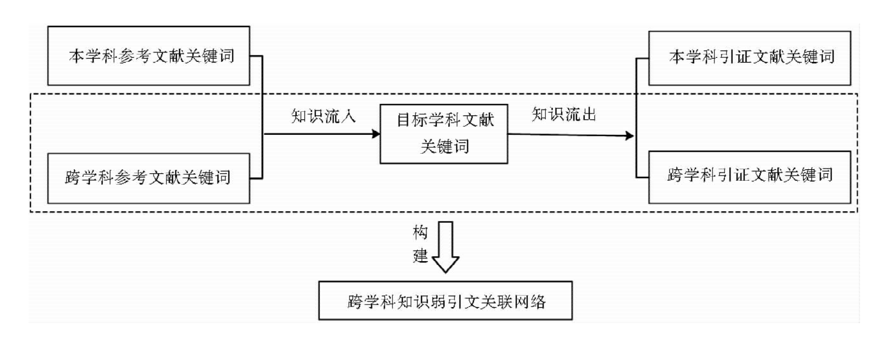


图2　基于关键词的跨学科知识弱引文关联网络构建

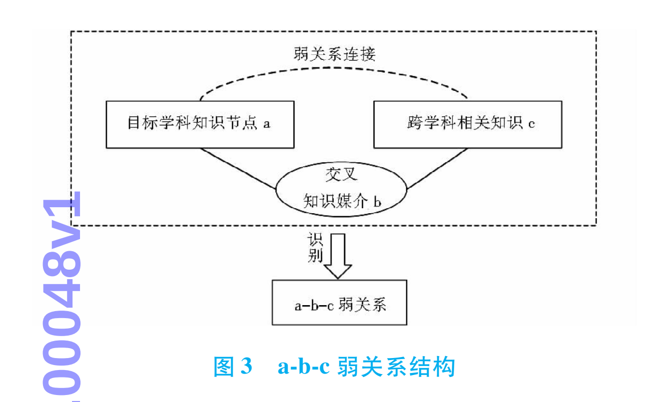


图3　a-b-c 弱关系结构

目标学科知识节点 a 与跨学科相关知识 c，通过知识媒介 b 建立联系。因为 a 与 b、b 与 c 有直接引文关系，但 a 与 c 没有，那么 a 与 c 就有潜在合作、共现于同一篇文章的可能。本文称引文知识关联网络中的这类 a-b-c 结构为弱关系结构。其中，节点 b 是使 a 与 c 通过弱关系建立连接的“关系桥”[2]，即知识媒介，在跨学科引文网络中，知识媒介往往是学科交叉节点。

2.2.1　知识媒介 b 类型划分

判断目标学科知识 a 能否与某跨学科相关知识 c 通过弱关系连接形成知识关联，首先寻找二者之间的知识媒介 b 作为“搭桥者”。跨学科文献的引用与被引用，本质上是将来自不同学科的知识基因进行自由组合，形成交叉知识进入到科学知识交流系统中，进而产生不同学科联系的知识链、知识网[21]。因此，学科交叉关键词是目标学科知识与跨学科相关知识建立弱连接的知识媒介。示意如图 4 所示：

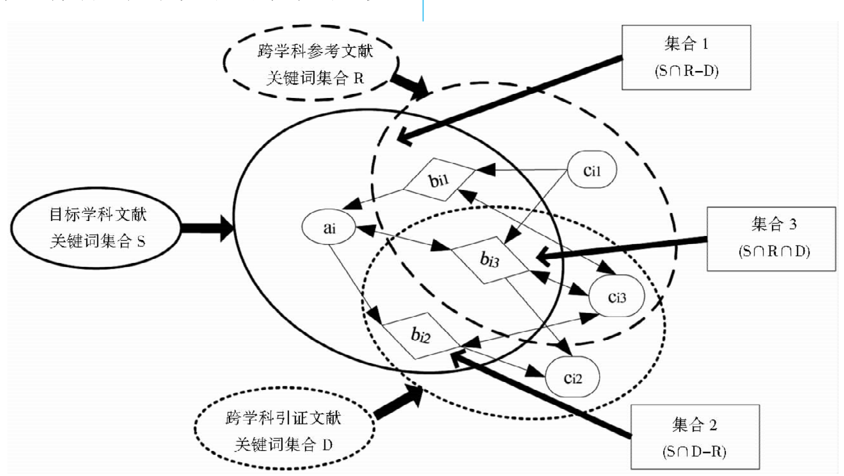


图4　知识媒介 b 类型与 a-b-c 弱关系连接路径

设目标学科文献关键词集合 S、其跨学科参考文献关键词集 R、引证文献关键词集合 D，则跨学科知识弱引文关联网络 G 是由 S、R、D 集合中的关键词，基于引用与被引用建立的知识关联。那么，学科知识节点 a 用关键词 a_i 表示，且 $a_i \in S$；跨学科相关知识 c 用关键词 c_i 表示，$c_i \in (R \cup D\text{-}S)$；知识媒介 b，用关键词 b_i 表示，$b_i \in$ 集合 1 ∪ 集合 2 ∪ 集合 3，其中集合 1($S \cap R\text{-}D$)、集合 2($S \cap D\text{-}R$)、集合 3($S \cap R \cap D$)表示网络 G 中跨学科交叉研究关键词。

根据知识流向，知识媒介 b_i 细分为 3 种类型：①流入型知识媒介 $b_{i1} \in$ 集合 1，跨学科知识经参考文献流入目标学科；②流出型知识媒介 $b_{i2} \in$ 集合 2，目标学科通过跨学科引证文献输出知识；③流动型知识媒介 $b_{i3} \in$ 集合 3，通过引用与被引在学科之间传递知识。

2.2.2　基于不同媒介类型的 a-b-c 弱关系流动路径识别

以学科知识节点 a_i 为例，其基于知识媒介 b_{1i}、b_{2i}、

b_{3i}，可分别与跨学科相关知识 c_i 建立连接，形成 a-b-c 弱关系结构。图 4 中，单向箭头表示参考或引证行为，即知识流入或流出；双向箭头表示互引行为，知识既有流入又有流出。则共有 7 种不同流动形式的弱连接路径，分别是：①基于"流入型"知识媒介 b_{i1} 的 2 种路径：$a_i \leftarrow b_{i1} \leftarrow c_{i1}$、$a_i \leftarrow b_{i1} \leftrightarrow c_{i3}$；②基于"流出型"知识媒介 b_{i2} 的 2 种路径：$a_i \rightarrow b_{i2} \leftrightarrow c_{i3}$、$a_i \rightarrow b_{i2} \rightarrow c_{i2}$；③基于"流动型"知识媒介 b_{i3} 的 3 种路径：$a_i \leftrightarrow b_{i3} \leftarrow c_{i1}$、$a_i \leftrightarrow b_{i3} \leftrightarrow c_{i3}$、$a_i \leftrightarrow b_{i3} \rightarrow c_{i2}$。编写程序，遍历网络 G 中的所有关系，抽取所有的 a-b-c 弱关系结构。

2.3 跨学科相关知识组合识别模型构建

学科知识节点 a_i 的活跃程度、知识媒介 bi 的中介能力、节点之间的联系强度，都是分析 a-b-c 弱关系结构中 a 与 c 潜在合作可能的影响性因素。因此，本文分别定义目标学科知识节点影响力指数 A_I、知识媒介影响力指数 B_I、跨学科知识相关性指数 C_I，对 a-b-c 弱关系各节点的特征及相关性进行量化描述；在此基础上，构建跨学科相关知识组合潜在合作指数 P 模型，以科学、合理地识别跨学科相关知识组合。

2.3.1 目标学科知识节点影响力指数 A_I

受热点浮现效应的影响，知识创新更易在呈明显上升趋势的热点主题研究过程中产生[22]。因此，研究热度呈明显上升趋势的学科知识节点，活跃度高，更有可能与跨学科知识进行合作。趋势分析法是一种经典的定量预测方法，其原理是从时间尺度上对关键词在不同时间点出现的频率进行回归分析，运用最小二乘法对历史数据拟合直线，分析变化率大小以预测其未来发展趋势[23]，有学者利用这一算法对学术名词的发展趋势进行判定[19, 24]。本文运用趋势分析法定义目标学科知识节点影响力指数 A_I，通过判断学科知识节点的热度变化趋势，描述其在学科研究中的活跃程度，公式为：

$$A_I = \frac{Y \times \sum_{y=1}^{Y}(y \times F_y) - \sum_{y=1}^{Y} y \times \sum_{y=1}^{Y} F_y}{Y \times \sum_{y=1}^{Y} y^2 - (\sum_{y=1}^{Y} y)^2} \quad 式(1)$$

式(1)中，Y 值表示数据样本的时间跨度(年数)，F_y 为关键词 a_i 第 y 年在目标学科的研究频次。A_I 是拟合直线的斜率(变化率)，若 $A_I > 0$，说明该关键词的研究热度为上升发展趋势，且 A_I 值越大，变化率越大，活跃度越高。

2.3.2 知识媒介影响力指数 B_I

根据弱关系理论，相比较于学科内部知识，跨学科知识之间的连接较难建立。知识媒介 b_i 作为关键搭桥者，与越多学科有关联，就越能接触到多样化的信息，激发知识创新的可能性就越大。因此，知识媒介的多学科程度是其媒介能力的重要体现。本文借鉴 A. L. Porter 等[16]提出的论文学科分布多样性测度指标，定义知识媒介影响力指数 B_I：

$$B_I = \frac{(F_1 + F_2 + \cdots + F_N)^2}{F_1^2 + F_2^2 + \cdots + F_N^2} = \frac{(\sum_{n=1}^{N} F_n)^2}{\sum_{n=1}^{N} F_n^2} \quad 式(2)$$

式(2)中，若知识媒介 b_i 在 N 个学科中出现过，则 F_n 表示第 n 个学科研究 b_i 的学术论文篇数，即 F_1、F_2 … F_N 之和为 b_i 在 N 个学科的总论文篇数。其中，$B_I \geqslant 1$，值越大，表明 b_i 跨学科多样性程度越高，影响力越大；当知识媒介 b_i 只在一个学科的文献中出现过时，B_I 值为 1。

2.3.3 跨学科知识相关性指数 C_I

该指标主要测度跨学科相关知识 c_i 通过知识媒介 b_i 与学科知识节点 a_i 建立弱关系连接的强度，体现了 a_i 与 c_i 的相关性程度。令 I_{ab} 为 a_i 与 b_i 的引用/被引用频次，I_{bc} 为 b_i 与 c_i 的引用/被引用频次，则节点 c_i 与 a_i 的相关性程度，与 I_{ab}、I_{bc} 呈正相关。基于此，本文根据前期研究文献[13]中定义的联系路径的可行性强度，定义跨学科相关知识相关性指数 C_I，计算公式为：

$$C_I = \frac{(I_{ab} \times I_{bc})^2}{|I_{ab} - I_{bc}| + \beta} \quad 式(3)$$

为使式(3)有意义，引进 β，当 $I_{ab} = I_{bc}$ 时，令 $\beta = 1$，当 $I_{ab} \neq I_{bc}$ 时，令 $\beta = 0$。

2.3.4 跨学科相关知识组合潜在合作指数 P

17 世纪，牛顿提出的万有引力定律是测量空间相互作用的一种方法[25]，其计算公式为：$I_{ij} = \frac{K \times M_i \times M_j}{D_{ij}}$，式中 I_{ij} 表示点 i 与 j 之间的引力大小；M_i、M_j 分别为点 i 和点 j 的质量，d_{ij} 为 i、j 之间的最短距离，k 为引力系数。

在跨学科相关知识组合中，学科知识节点 a_i 通过知识媒介 b_i 与跨学科相关知识 c_i 产生合作可能性的大小，可视为在整个引文关联网络中 a_i、c_i 节点引力的大小，受学科关键知识节点影响力指数 A_I、知识媒介影响力指数 B_I、跨学科知识相关性指数 C_I 的影响。因此，基于引力模型定义跨学科相关知识组合潜在合作指数 P，其公式为：

$$P = A_I \times B_I \times C_I = \frac{Y \times \sum_{y=1}^{Y}(y \times F_y) - \sum_{y=1}^{Y} y \times \sum_{y=1}^{Y} F_y}{Y \times \sum_{y=1}^{Y} y^2 - (\sum_{y=1}^{Y} y)^2} \times \frac{(\sum_{n=1}^{N} F_n)^2}{\sum_{n=1}^{N} F_n^2} \times \frac{(I_{ab} \times I_{bc})^2}{|I_{ab} - I_{bc}| + \beta} \quad 式(4)$$

3 实证分析——情报学跨学科相关知识组合识别

3.1 数据来源与预处理

本文选择CSSCI来源期刊中9种情报学期刊刊载的论文作为实证样本进行研究，验证基于弱引文关系跨学科知识合作组合识别方法的可行性与有效性。9种期刊包括:《情报学报》《情报资料工作》《图书情报工作》《情报理论与实践》《图书情报知识》《图书与情报》《情报科学》《数据分析与知识发现》《情报杂志》。时间跨度为2015－2019年。

(1)源文献数据下载。利用CNKI数据库，通过专业检索功能，检索上述9种期刊5年间刊载的有效文献共18 052篇，下载含标题、关键词、摘要、出版时间等字段的题录信息，批量下载为test文本格式，并导入MySQL关系型数据库中存储。

(2)参考/引证文献数据获取。运用编码的Python爬虫程序，获取18 052篇论文的参考文献和引证文献的题录数据，包括标题、关键词、摘要、出版时间、刊名等。

(3)跨学科参考/引证文献匹配与预处理。根据《中国科技期刊引证报告》中对期刊－学科的分类，用文献[26]的方法，读取题录数据中的“刊名”字段信息，并匹配判断文献的学科归属。去除文献引用格式错误、缺失等无效文献和英文文献，筛选来自跨学科中文期刊的参考和引证文献数据，结果同样下载为test文本格式，导入MySQL中，并分别对应其目标学科数据信息。最终得45 086篇跨学科参考文献和40 103篇跨学科引证文献。

3.2 构建情报学跨学科知识弱引文关联网络

根据2.1的方法描述，对经过预处理的样本数据，构建情报学科文献关键词－跨学科参考文献关键词－跨学科引证文献关键词的知识关联网络，操作步骤如下:

(1)提取关键词。因为论文自带关键词数量有限，不能很好地反映全文内容，本文抽取MySQL数据库中存储的目标学科文献及其对应的跨学科参考和引证文献“标题”和“摘要”字段的数据，构成语料库。利用MySQL关系型数据库中的“text mining”选项，编写Python语言工具，应用Jieba分词包，将中国规范术语数据库提供的各学科规范关键词作为分词词典，对语料库进行中文分词、去停用词、词频统计、合并同义词、统一规范化等处理，不考虑频次，只要出现就保存，得到代表每篇文献核心知识点的关键词，并将抽取到的关键词对应其所属的文献保存。经过最终处理，得到21 023个情报学科关键词的集合S、52 805个跨学科参考文献关键词的集合R、51 917个跨学科引证文献关键词的集合D。

(2)构建引文关系网络。利用MySQL中的关系型规则匹配方法，将基于文献的引文关系转换为所包含关键词之间的引证关系。同时，编写Python程序，遍历S、R、D集合中的所有关键词及其引证关系，构建基于关键词的跨学科知识引文关系网络G。该网络是有向加权网络，反映关键词之间的引证关系，方向表示引用或被引用，权重代表两者引用/被引用行为产生的连接次数。

3.3 识别情报学跨学科引文网络中的a-b-c弱关系结构

依据2.2方法论述，在情报学跨学科引文关系网络G中，识别知识媒介b和a-b-c弱关系结构，方法与步骤如下:

(1)知识媒介b识别。学科交叉研究关键词是a-b-c弱关系结构建立的重要知识媒介，因此首先需提取学科交叉研究关键词。如图4所示，对集合S、集合R、集合D中的关键词进行共现分析，得到3部分交叉研究关键词，集合1中有4 817个流入型知识媒介、集合2中2 805个流出型知识媒介、集合3中218个流动型知识媒介。

(2)a-b-c弱关系结构识别。以S集合中情报学的每一个关键词为起点，利用Apriori关联游走算法，编写Python程序，遍历MySQL数据库及其所有的关联数据，寻找S集合中的学科知识节点a在网络G中的知识媒介b、7种不同连接路径的所有跨学科相关知识c，去除a与c的直接引用与被引用关系，最终仅得到657条a-b-c弱关系结构。

3.4 情报学跨学科相关知识组合指数计算与识别

(1)相关指数计算。对表征情报学科知识节点的所有关键词a，抽取MySQL数据库中存储记录的“出版时间”字段信息，统计该关键词在2015－2019年内每一年的出现频数，应用公式(1)，计算知识节点影响力指数A_1，计算结果见表1第6列;对表征知识媒介b的所有关键词，利用MySQL数据库获取其所属的全部文献记录，根据期刊－学科分类来确定关键词分布在哪些学科，并分别统计于每一个学科的分布篇数，应用公式(2)，得知识媒介影响力指数B_1，计算结果见表1第

图书情报工作
第 64 卷 第 21 期 2020 年 11 月

7 列;利用自编的 Python 程序,遍历网络 G,得每一条 a-b-c 弱关系结构中 b 与 a、b 节点的连接次数,应用公式(3),计算跨学科知识相关性指数 C_I,见表 1 第 8 列;最后,应用公式(4),计算每一条 a-b-c 弱连接的 a 与 c 潜在合作指数 P,见表 1 第 9 列。

(2)跨学科相关知识组合识别结果。按照跨学科相关知识 a 与 c 潜在合作指数 P 值由高到低顺序排列,选择排名前 10 位的结果展示,相关数据如表 1 所示:

表 1 情报学科跨学科相关知识组合识别结果相关数据(部分)

排名	目标学科知识节点 a	知识媒介 b	跨学科相关知识 c	流动路径	目标学科知识节点影响力指数 A_I	知识媒介影响力指数 B_I	跨学科知识相关性指数 C_I	潜在合作指数 P
1	科研合作	知识流动	种群动力学模型	a↔b←c	2.60	2.73	5.33	37.35
2	引文网络	多元数据融合	有限元分析法	a←b←c	2.50	2.34	6.25	35.85
3	跨学科	链路预测	随机森林算法	a←b↔c	2.60	1.99	6.25	32.48
4	高被引文献	复杂网络	羊群效应	a←b←c	2.70	2.83	4.00	30.87
5	智库	大数据	MongoDB 数据库	a↔b→c	2.30	3.24	4.00	29.82
6	Altmetrics	开放存取	信息网络传播权	a↔b→c	3.10	2.08	4.50	29.25
7	知识图谱	关联规则	多标签学习	a↔b←c	2.00	2.17	6.25	26.77
8	网络舆情	实体相似度	信息茧房	a↔b→c	2.50	2.09	4.00	20.60
9	睡美人文献	超弦理论	超弦引力模型	a←b→c	2.40	1.68	4.50	18.37
10	社交媒体	人际情报网络	概念格	a→b←c	2.30	1.73	4.00	16.02
…	…	…	…	…	…	…	…	…

4 识别结果分析

4.1 识别结果有效性分析

4.1.1 目标学科知识节点影响力指数 A_I 可有效识别呈上升趋势的学科热点

表 1 中,第 2 列数据是表征目标学科知识节点 a 的关键词,其中"Altmetrics""智库""科研合作""跨学科"等的第 6 列影响力指数 A_I 值相对较高,说明情报学科 2015－2019 年间这些关键词的研究文献呈逐年增多的趋势,是学科研究热点。这些结果恰恰是文献[27]识别得到的情报学领域发展趋势为"上升类"研究热点的一部分。"睡美人文献""高被引文献"等也是近年情报学中研究热度与关注度持续增长的关键词[28－30]。这说明,基于趋势分析法的 A_I 指数用于分析上升趋势的研究热点是可行有效的,识别结果有更大可能引入其他学科的相关概念、理论、方法和技术等跨学科相关知识实现跨学科合作。

4.1.2 知识媒介影响力指数 B_I 可识别跨学科性强的搭桥者

观察表 1 第 3 列和第 7 列数据,"大数据"的知识媒介影响力指数 B_I 为 3.24,数值最高。表明其跨学科性明显,涉及学科领域多样,相关知识广泛,具有较高的媒介能力,成为促进不同学科间知识交流与融合的有力"搭桥者"。同时,研究表明:大数据作为信息社会纵深发展的产物,其研究已经成为科技界的研究热点,横跨信息科学、社会科学、数学、教育学、心理学、经济学等诸多学科领域,表现出典型的跨学科性[31－32],这与本文的研究结论存在一致性。因此,知识媒介影响力指数 B_I 可用来计算关键词的跨学科性,识别媒介能力强的"搭桥者"。

4.1.3 跨学科知识相关性指数 C_I 能有效识别与目标学科相关性高的跨学科知识

表 1 中,第 4 列和第 8 列数据,关键词"有限元分析法""随机森林算法""多标签学习""种群动力学模型"等的跨学科知识相关性指数 C_I 值较高,说明它们分别基于知识媒介"多元数据融合""链路预测""关联规则""知识流动"作为"搭桥者",与情报学科的"引文网络""跨学科""知识图谱""科研合作"等关联度高。例如,引文网络－多元数据融合－有限元分析法,流动路径为 a←b←c,"引文网络"引用"多元数据融合"频次高,"多元数据融合"引用"有限元分析法"频繁,说明"有限元分析法"可以用于"引文网络"的研究。文献分析发现,二者有一定的关联性,见 4.2 的分析结论。因此,相关性指数 C_I 可以有效识别与目标学科相关性高的跨学科知识。

4.1.4 潜在合作指数 P 能有效结合 a-b-c 弱连接中目标学科知识节点 a 的活跃度、知识媒介 b 的跨学科性、跨学科相关知识 c 的相关性特征,识别具有高合作潜力的跨学科相关知识组合

表 1 中,观察第 9 列潜在合作指数 P 值排名第一

位的“科研合作－知识流动－种群动力学模型”a-b-c弱关系的相关数据，其中 A_1 值为2.60，BI 值为2.73，CI 值为5.33，分别位居第4位、第3位、第2位，虽然目标学科知识节点活跃度、知识媒介多样性程度、跨学科知识相关性不是最高，但都比较高，所以“科研合作”与“种群动力学模型”的吸引力最强，之间有比较密切的潜在联系，弱关系连接的潜力值最大，为37.35，成为合作潜力最高的跨学科相关知识组合。因此，潜在合作指数 P 能有效结合 a-b-c 弱关系中各节点的属性及相互关系，识别具有高合作潜力的跨学科相关知识组合。

4.2 识别结果应用分析

表1为识别得到合作潜力值高的跨学科相关知识组合。为探讨识别结果的有效性，分析跨学科相关知识解决情报学问题的技术方案，促进跨学科合作研究创新，对排名前两位的结果进行应用分析：

（1）“科研合作”↔“知识流动”←“种群动力学模型”。“科研合作”可以使不同的知识实现集成，使不同知识背景的研究人员、组织或机构之间进行知识的碰撞，是加速知识扩散的重要方式之一[33]。从科研合作角度出发，合作双方基于知识势能差获取互补性知识资源实现知识流动。“知识流动”过程推动知识传播、扩散、融合及创新，形成知识链、知识网[19]。“种群动力学模型”是一种研究种群间以及种群与不确定性环境间相互作用的行为动力学模型[34]。知识流动与种群移动有一定的知识关联，知识流动又嵌入在科研合作过程中。因此，可尝试将种群动力学模型代入科研合作过程，模拟分析该过程中知识扩散模式、知识融合的演化路径等，挖掘科研合作中主体合作模式、关系结构、地位变化等特征，以揭示科研合作发展的潜在规律。

（2）“引文网络”←“多元数据融合”←“有限元分析法”。“引文网络”是由研究文献构成的大规模知识网络，包含了大量的引用关系和文本属性等多元数据。“多元数据融合”是指利用数据融合算法高效整合多种关联数据，从而通过更丰富的信息抽取与识别，获得更加准确的知识单元之间的潜在语义关系[35]。“有限元分析法”将求解域看成是由许多称为有限元的小的互连子域组成，对每一单元假定一个合适的近似解，然后推导求解这个域总的满足条件，从而得到问题的解[36]。引文网络中存在各种各样的复杂网络，按照有限元分析法的思想，可将这些复杂网络分解为多个简单的子网络，对每个子网络中的多元数据进行有效融合，用较简单的问题代替复杂问题，继而模拟得到实现多元数据融合的整体网络。因此，可以尝试应用有限元分析法，融合引文网络中的多元数据，提高引文网络数据挖掘和知识发现的效率。

5 结语

本文基于弱引文关系提出跨学科相关知识组合发现方法。首先，构建基于关键词的跨学科知识弱引文关联网络；其次，识别引文网络中 a-b-c 弱关系结构，定义学科关键知识节点影响力指数 A_1、知识媒介影响力指数 B_1、跨学科知识相关性指数 C_1，构建合作潜力指数 P 模型。最后，选择情报学领域9种优秀期刊2015－2019年的载文及其跨学科参考/引证文献为样本进行实证研究。对识别结果进行有效性分析，发现本文提出的跨学科相关知识组合发现方法具有合理性和有效性；同时，探讨识别结果的应用前景，发现基于该方法的识别结果具有一定的跨学科合作意义。

本研究的实证样本来源于情报学科的中文文献。一方面，该方法是否适用于其他学科或其他语种文献，有待于进一步验证；另一方面，网络社交媒体等非正式学术交流方式的发展，也为跨学科知识的交流与合作提供了研究视角与数据来源。因此，未来研究可利用跨语言文献及网络社交媒体数据进一步对跨学科相关知识组合的识别工作展开研究，以得到更深入、全面的结论，从而进一步促进学科间知识的交流合作，推动知识创新。

参考文献：

[1] GATES A J, KE Q, VAROL O, et al. Nature's reach: narrow work has broad impact[J]. Nature, 2019, 575(7781): 32－34.
[2] GRANOVETTER M. The strength of weak ties[J]. The American journal of sociology, 1973, 78(6): 1360－1380.
[3] ONNELA J P, SARAMAKAI J, HYVONEN J, et al. Structure and tie strengths in mobile communication networks[J]. Proceedings of the National Academy of Sciences of the United States of America, 2007, 104(18): 7332－7336.
[4] BAKSHY E, ROSENN I, MARLOW C, et al. The role of social networks in information diffusion[C]/Proceedings of the 21st annual conference on World Wide Web. Lyon: Annual conference on World Wide Web, 2012: 519－528.
[5] ZHAO J, WU J, XU K. Weak ties: subtle role of information diffusion in online social networks[J]. Physical review E, 2010, 82(1): 16105.
[6] DAVID E, JON K. Networks, crowds, and markets: reasoning about a highly connected world[M]. United Kingdom: Cambridge University Press, 2012.

[7] GENIUS S K. Published literature and diffusion of medical innovation: exploring innovationgeneration[J]. The Canadian journal of information and library science,2005, 29(1): 27 - 54.
[8] ABBASI A, ALTMANNI J, HOSSAIN L. Identifying the effects of co-authorship networks on the performance of scholars: a correlation and regression analysis of performance measure and social network analysis measures[J]. Journal of informetrics, 2011, 5(4): 594 - 607.
[9] BETTONI M, SCHILLER G, BERNHARD W. Weak ties cooperation in the CoRe knowledge network [C]//The 9th European conference on knowledge management. Southampton: Southampton Solent University,2008:1 - 7.
[10] YANG L Y, MORRIS S. Mapping institutions and their weak ties in a specialty: a case stud y of cystic fibrosis body composition research[J]. Scientometrics, 2009, 79(2): 421 - 434.
[11] 隗玲,许海云,郭婷,等. 基于弱共现和突发监测的情报学学科研究主题及交叉性分析[J]. 图书情报工作,2015, 59(21): 105 - 114.
[12] 李长玲,刘小慧,刘运梅,等. 基于开放式非相关知识发现的潜在跨学科合作研究主题识别——以情报学与计算机科学为例[J]. 情报理论与实践,2018, 41(2): 100 - 104.
[13] 刘小慧,李长玲,崔斌,等. 基于闭合式非相关知识发现的潜在跨学科合作研究主题识别——以情报学与计算机科学为例[J]. 情报理论与实践, 2017, 40(9): 71 - 76.
[14] SONG M, KANG K, YOUNG A. Investigating drug-disease interactions in drug-symptom-disease triples via citation relations [J]. Journal of the Association for Information Science and Technology, 2018, 69(11): 1355 - 1368.
[15] 梁镇涛,巴志超,徐健. 基于引文的跨学科领域发展路径分析——以眼动追踪领域为例[J]. 图书情报工作, 2019, 63(23): 65 - 78.
[16] PORTER A L, COHEN A S, ROESSNER D J, et al. Measuring researcher interdisciplinarity[J]. Scientometrics, 2007, 72(1): 117 - 147.
[17] 林向义,罗洪云,李秀成. 企业个体从社交媒体网络吸收异质性知识的过程机理:弱连接关系视角[J]. 情报理论与实践, 2019, 42(3): 65 - 71.
[18] 杜德慧,李长玲,相富钟,等. 基于引文关键词的跨学科相关知识发现方法探讨[J/OL]. 情报杂志:1 - 6[2020 - 09 - 21]. http://kns. cnki. net/kcms/detail/61. 1167. G3. 20200721. 1527. 004. html.
[19] 张瑞,赵栋祥,唐旭丽,等. 知识流动视角下学术名词的跨学科迁移与发展研究[J]. 情报理论与实践,2020, 43(1): 47 - 55.
[20] LOUADI M, MOHAMED El. Knowledge heterogeneity and social network analysis-towards conceptual and measurement clarifications [J]. Knowledge management research & practice, 2008, 6(3): 199 - 213.
[21] 王雪,李睿. 知识生态学视角下的各类引文现象阐释[J]. 情报杂志, 2018, 37(9): 179 - 184
[22] FAJARDO-ORTIZ D, LOPEZ-CERVANTES, DURAN L, et al. The emergence and evolution of the research fronts in HIV/AIDS research[J]. Plos one, 2017, 12(5): e178293.
[23] 周复恭, 黄运成. 应用线性回归分析[M]. 北京:中国人民大学出版社, 1989.
[24] 李峰,刘静延,蒋录全. 预测方法的发展及最新动态[J]. 情报杂志, 2005(6): 76 - 77.
[25] 牛顿. 自然哲学的数学原理[M]. 赵振江,译. 北京:商务印书馆, 2009.
[26] 冯志刚,李长玲,刘小慧,等. 基于引用与被引用文献信息的图书情报学跨学科性分析[J]. 情报科学,2018, 36(3): 105 - 111.
[27] 李长玲,牌艳欣,相富钟,等. 改进 z 指数的高被引学科研究热点识别方法探讨[J]. 情报理论与实践,2020(6): 39 - 45,66.
[28] 李月琳,章小童,王姗姗,等. 情报学的坚守与拓展——基于 2018 年 ASIS&T 年会论文的综述[J]. 图书情报知识, 2019(3): 4 - 16.
[29] 刘运梅,李长玲,杜德慧. 时间序列视角下 PLOS ALM 指标特性识别模型构建与应用[J]. 情报资料工作, 2019, 40(6): 6 - 15.
[30] 宗张建. 睡美人文献识别方法研究进展[J]. 图书情报工作, 2019, 63(16): 132 - 142.
[31] 吕晓赞,王晖,周萍. 中美大数据论文的跨学科性比较研究[J]. 科研管理, 2019, 40(4): 1 - 13.
[32] 李国杰,程学旗. 大数据研究:未来科技及经济社会发展的重大战略领域——大数据的研究现状与科学思考[J]. 中国科学院院刊,2012, 27(6): 647 - 657.
[33] 李纲,巴志超. 科研合作超网络下的知识扩散演化模型研究[J]. 情报学报, 2017, 36(3): 274 - 284.
[34] KŘIVAN V, LEWIS M, BENTZ B J, et al. A dynamical model for bark beetle outbreaks[J]. Journal of theoretical biology, 2016, 407(21): 25 - 37.
[35] 陈文杰,许海云. 一种基于多元数据融合的引文网络知识表示方法[J]. 情报理论与实践, 2020, 43(1): 150 - 154.
[36] BREKELMANS A M, POORT H W, SLOOFF J H. A new method to analyse the mechanical behaviour of skeletal parts[J]. Acta orthopaedica scandinavica, 2009, 43(5): 301 - 317.

作者贡献说明:

牌艳欣:数据搜集与处理,论文撰写与修改;
李长玲:研究设计与指导,论文撰写与修改;
徐璐:论文修改与校对。

Discussion on the Method of Interdisciplinary Related Knowledge Combination Identification on the Perspective of Weak Citation Relationship ——Taking Information Science for Example

Pai Yanxin Li Changling Xu Lu

Science and Technology Information Research Institute, Shandong University of Technology, Zibo 255049

Abstract: [**Purpose/significance**] With the complexity of the scientific system, interdisciplinary research has become an important paradigm and inevitable trend of modern scientific innovation research. Identifying interdisciplinary relevant knowledge combination that has high cooperation potential, becomes the key to promoting interdisciplinary scientific research cooperation and innovation. [**Method/process**] Firstly, the paper selected target subject source literature, its interdisciplinary reference literature, and its interdisciplinary citing literature, so as to construct an interdisciplinary knowledge weak reference relational network based on keywords. Secondly, it classified the types of Knowledge Medium b, and identified the weak relational structure of the Knowledge Node a of the target discipline-Knowledge Medium b - Interdisciplinary Knowledge c. Finally, the paper defined the Knowledge Node Influence Index AI of the target discipline, the Knowledge Media Influence Index BI, the Interdisciplinary Knowledge Correlation Index CI, the Interdisciplinary Knowledge Combination a-c and Potential Cooperation Index P, to identify the interdisciplinary knowledge combination with high cooperation potential. [**Result/conclusion**] The papers of 9 CSSCI journals in the field of informatics from 2015 to 2019 and their interdisciplinary references and citations were selected as samples for empirical research, to verify the effectiveness and feasibility of the discovery method of interdisciplinary knowledge combination based on weak citation relationship, and to identify the interdisciplinary knowledge combination with high cooperation potential of intelligence disciplines such as "scientific research cooperation" - "knowledge flow" - "population dynamics model".

Keywords: strength of weak ties citation analysis a-b-c weak connection interdisciplinary related knowledge combination potential cooperation index P